Multi-Core Microprocessor Chips: Motivation & Challenges

Dileep Bhandarkar, Ph. D.
Architect at Large
DEG Architecture & Planning
Digital Enterprise Group
Intel Corporation

October 2005

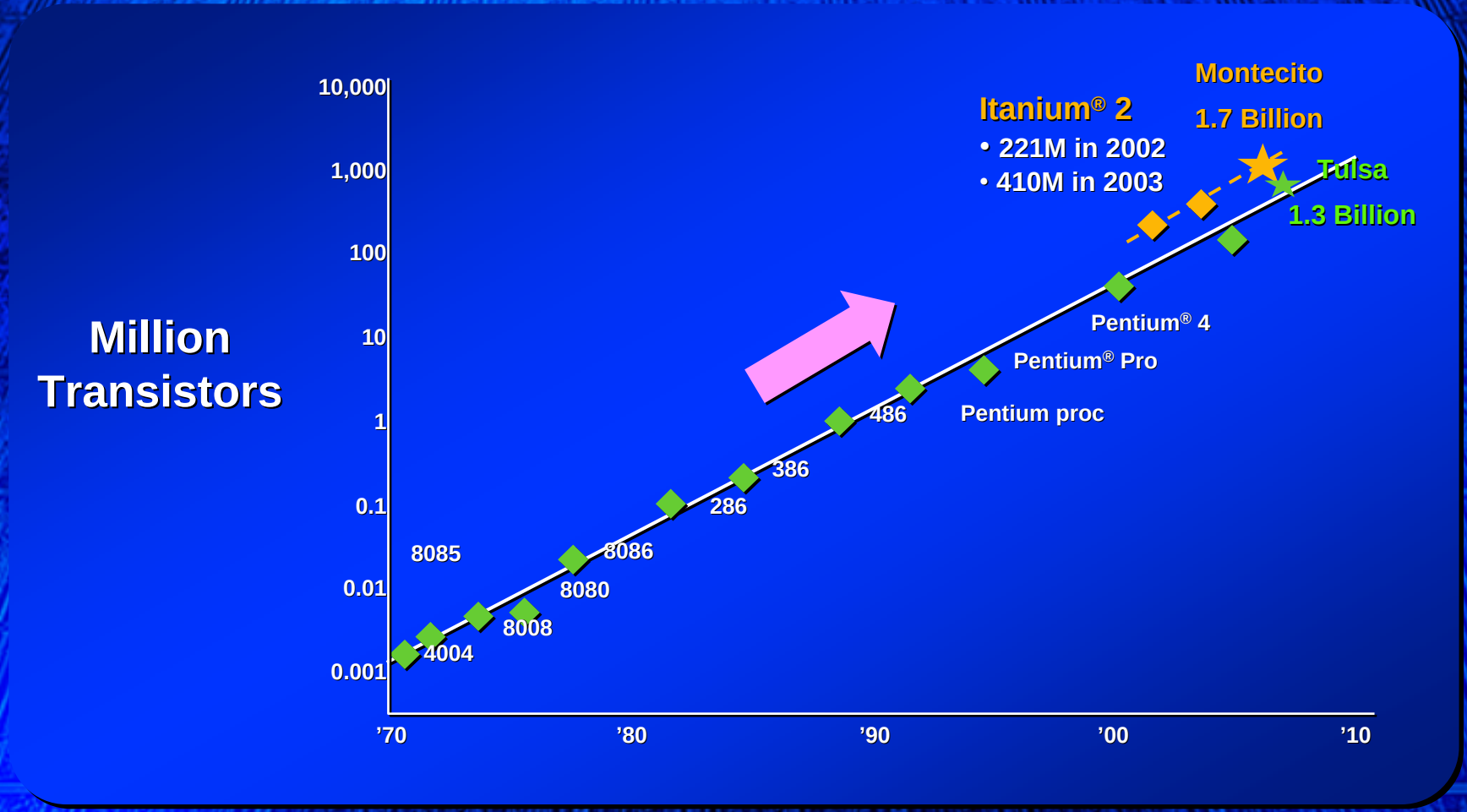
Copyright © 2005 Intel Corporation.

Agenda

- Semiconductor Technology Evolution
- Design Challenges
- Why Multi-Core Processor Chips?
- Power/Performance Trade-Offs
- CMP Directions
- Beyond CMP
- Summary

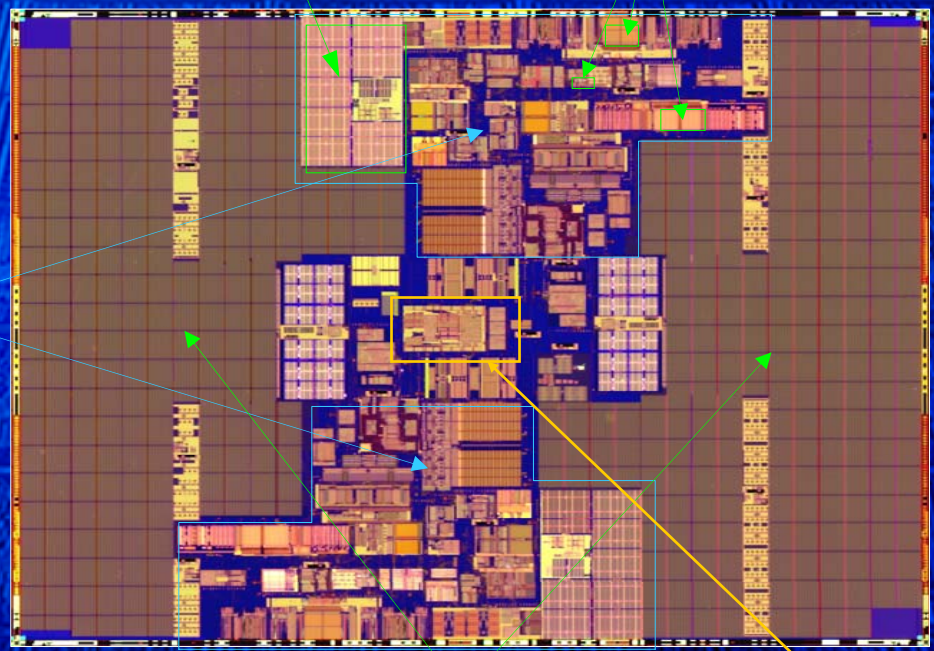
©2005, Intel Corporation
Intel, the Intel logo, Pentium, Itanium and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries
*Other names and brands may be claimed as the property of others

40 Years of Moore's Law



More than 1 Billion Transistors in 2005!

First Billion Transistor Dual Core Chip



1.72 Billion Transistors

Continuation of Moore's Law

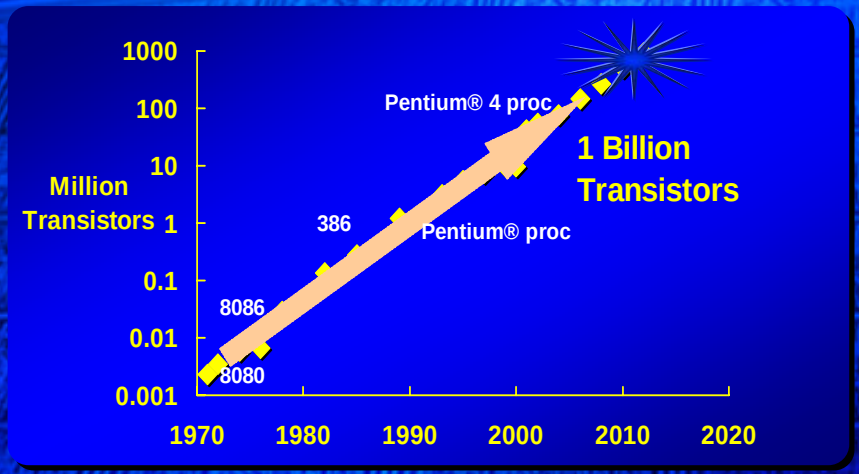
Process Name	P856	P858	Px60	P1262	P1264	P1266	P1268	P1270
1st Production	1997	1999	2001	2003	2005	2007	2009	2011
Process Generation	0.25μm	0.18μm	0.13μm	90 nm	65 nm	45 nm	32 nm	22 nm
Wafer Size (mm)	200	200	200/300	300	300	300	300	300
Inter-connect	Al	Al	Cu	Cu	Cu	Cu	Cu	?
Metal layers	5	6	6	7	8	8-9?	9-10?	10?
Channel	Si	Si	Si	Strained Si	Strained Si	Strained Si	Strained Si	Strained Si
Gate dielectric	SiO_2	SiO_2	SiO_2	SiO_2	SiO_2	High-k	High-k	High-k
Gate electrode	Poly-silicon	Poly-silicon	Poly-silicon	Poly-silicon	Poly-silicon	Metal	Metal	Metal
Lithography	248 nm	248 nm	248 nm	193 nm	193 nm	193 nm	EUV 13.4 nm	EUV 13.4 nm

Subject to change

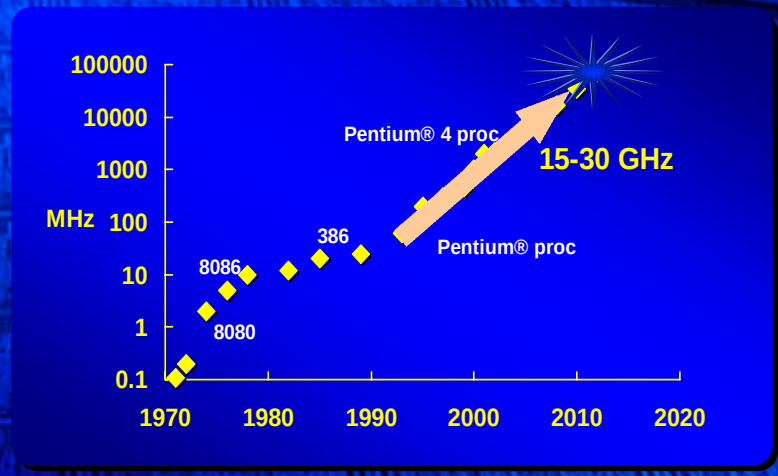
intel

Source: Intel

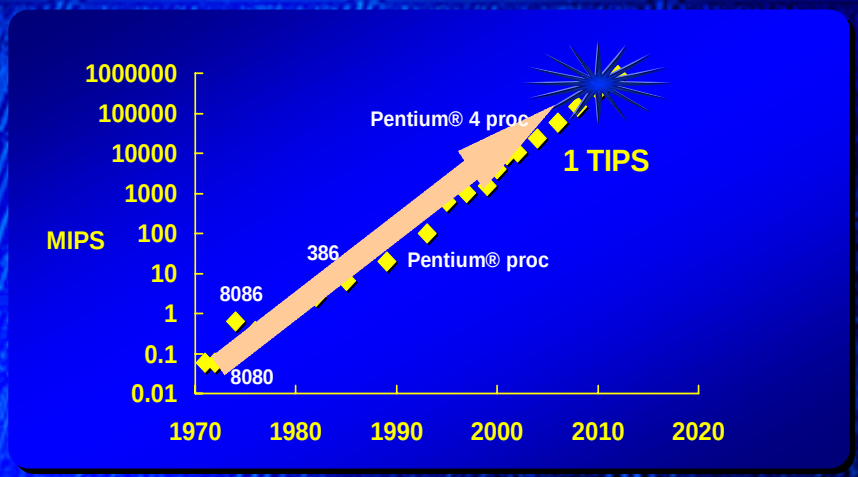
Power Will be the Limiter



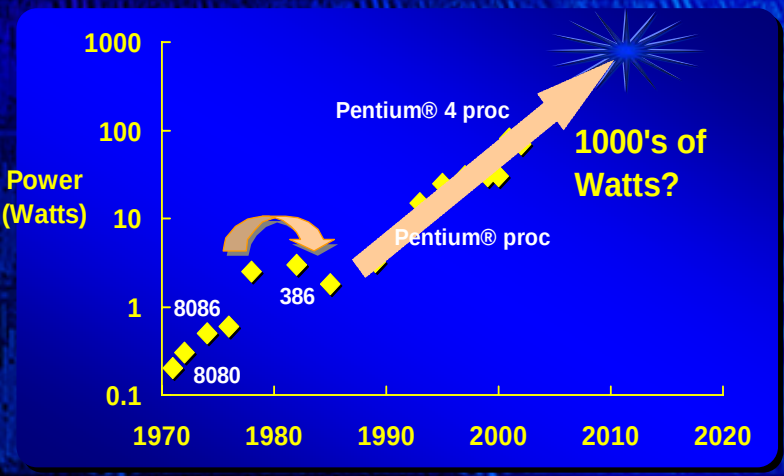
>1B transistor integration capacity *will* exist



Clock frequency can increase



Applications will demand TIPS performance



But the Power…

Challenge: Highest performance in the power envelope

Agenda

- Semiconductor Technology Evolution
- Design Challenges
- Why Multi-Core Processor Chips?
- Power/Performance Trade-Offs
- CMP Directions
- Beyond CMP
- Summary

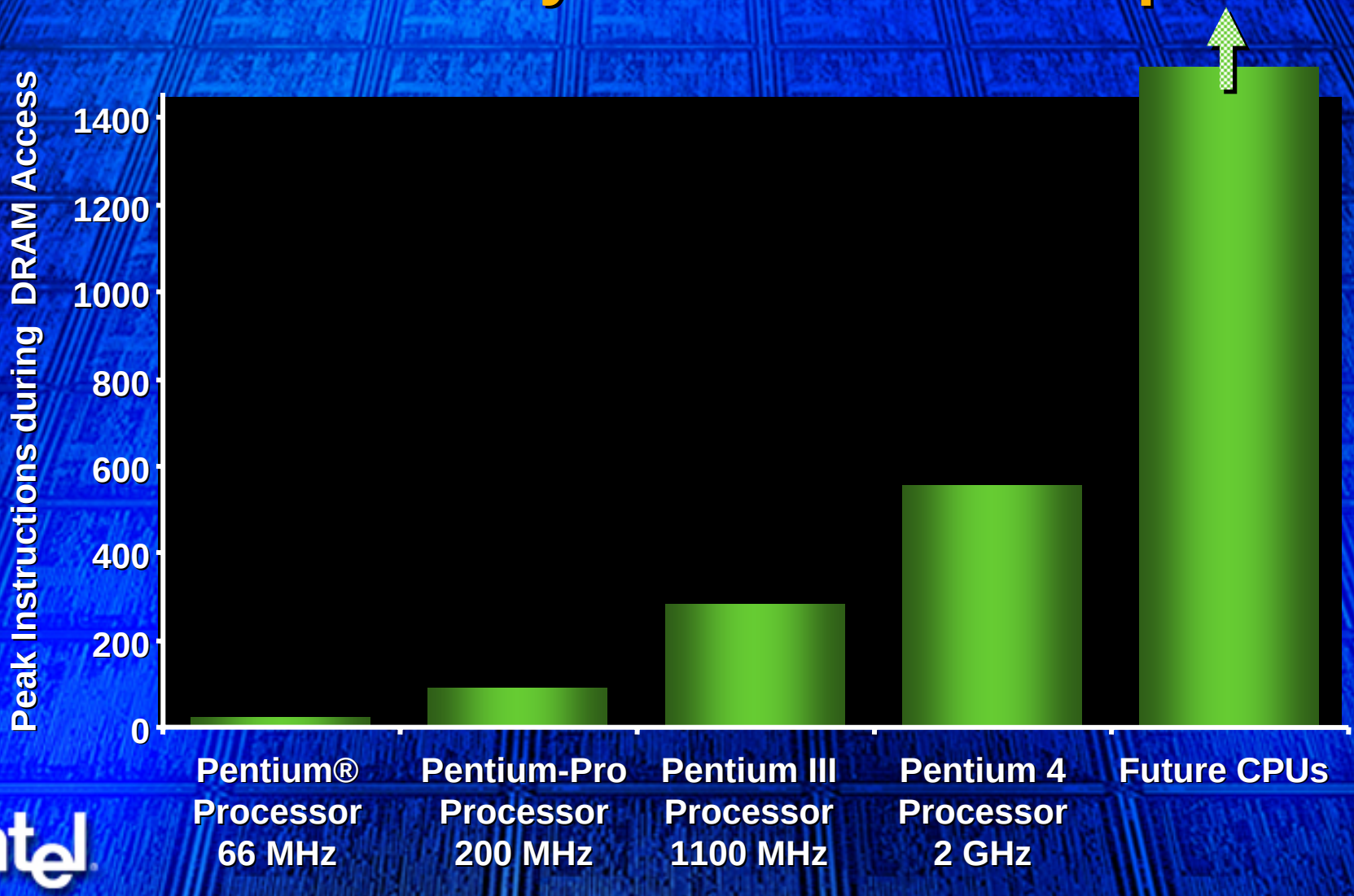
©2005, Intel Corporation
Intel, the Intel logo, Pentium, Itanium and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries
*Other names and brands may be claimed as the property of others

Design Challenges

- Memory latency not scaling as fast as processor speed
- Power growing non-linearly with single thread performance
- Designer productivity lagging design complexity
- Ability to validate and test complex design
- Keeping up with new process technology every two years

intel

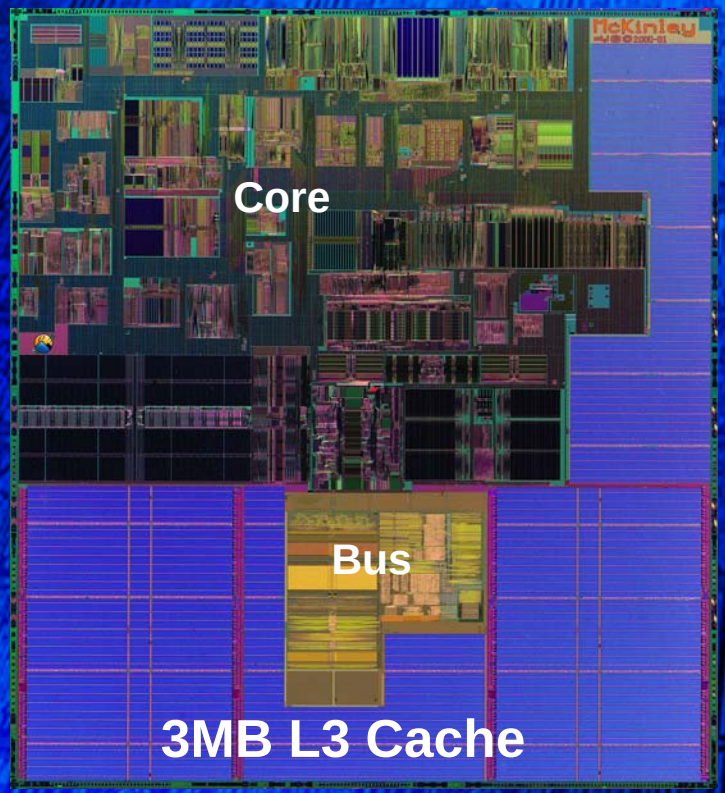
Long Latency DRAM Accesses: Needs Latency Tolerant Techniques



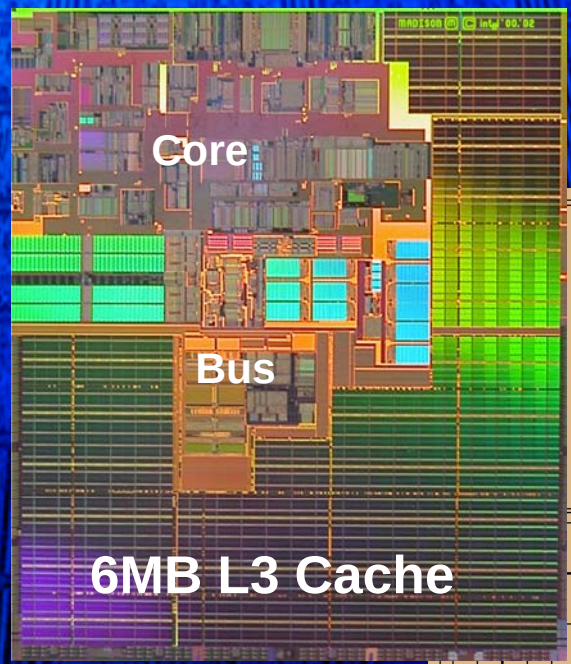
DRAM Latency Tolerance

- Continue building even larger caches
  - Every semiconductor process generation provides opportunity to double cache size
  - Cache becomes larger part of die
- Hide multiple threads of execution behind memory latency
- Intel implemented simultaneous multi-threading in 2000
- Implement multi-core products as Moore's Law allows

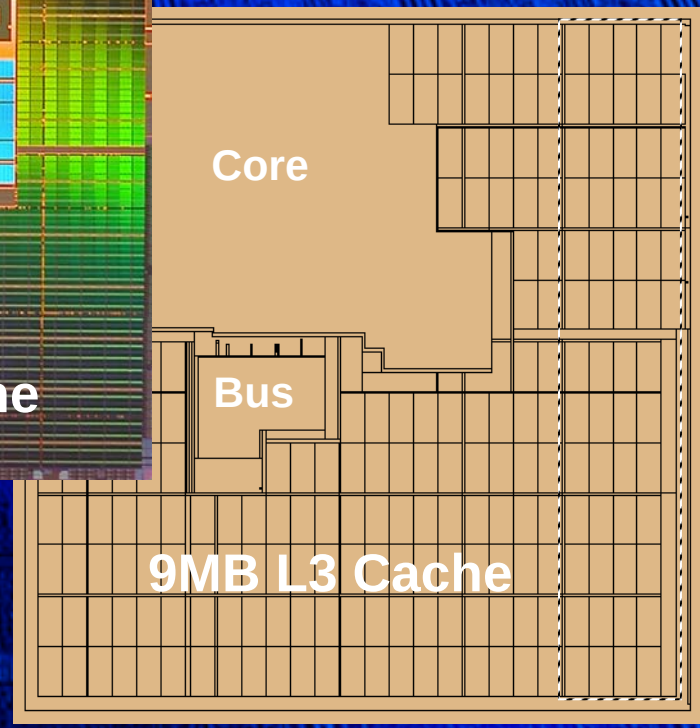
From 180 nm to 130 nm



180 nm



130 nm



Cache becomes a larger part of the Die Area.

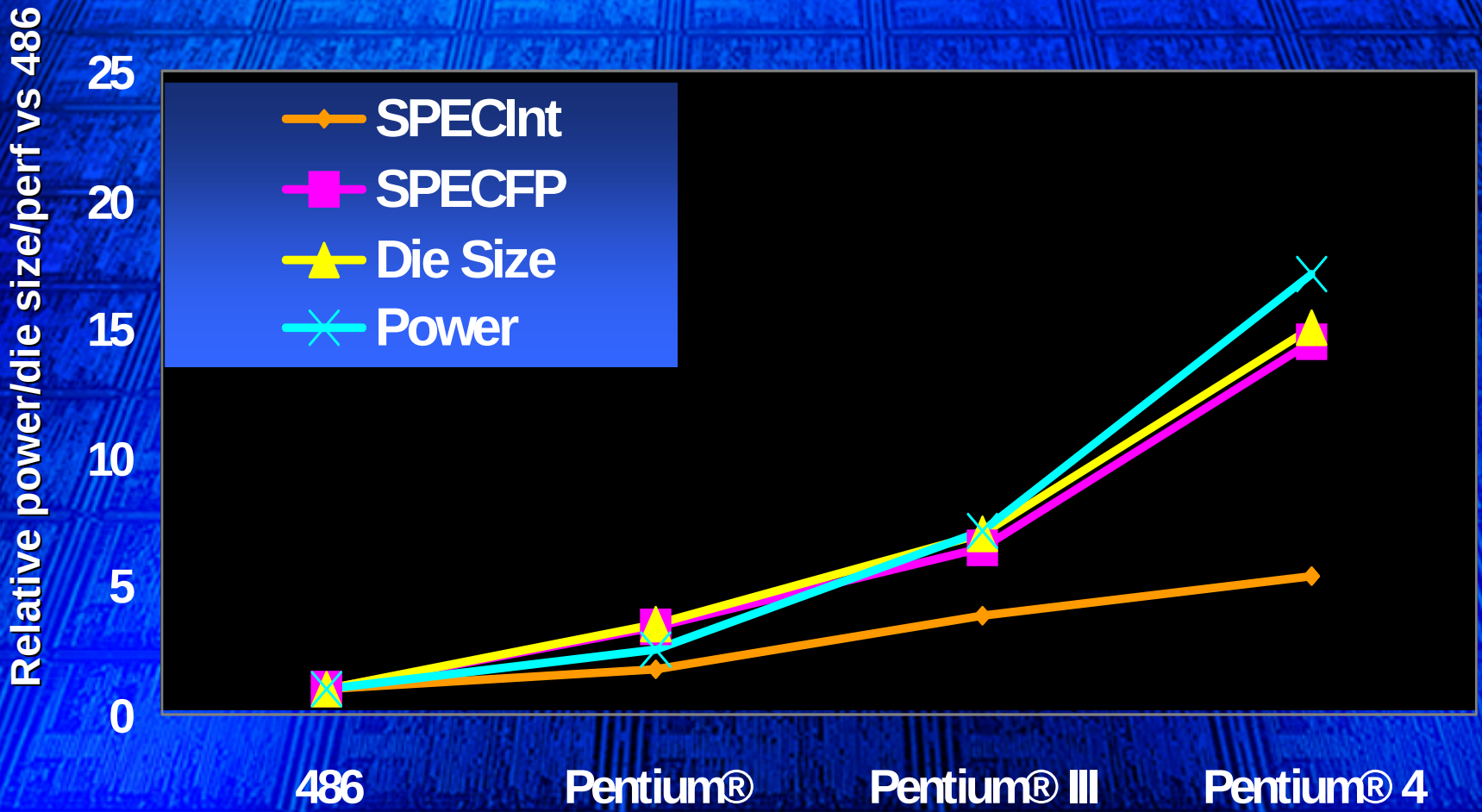
Dual core does not double die size.

Agenda

- Semiconductor Technology Evolution
- Design Challenges
- Why Multi-Core Processor Chips?
- Power/Performance Trade-Offs
- CMP Directions
- Beyond CMP
- Summary

©2005, Intel Corporation
Intel, the Intel logo, Pentium, Itanium and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries
*Other names and brands may be claimed as the property of others

Single-stream Performance vs Costs



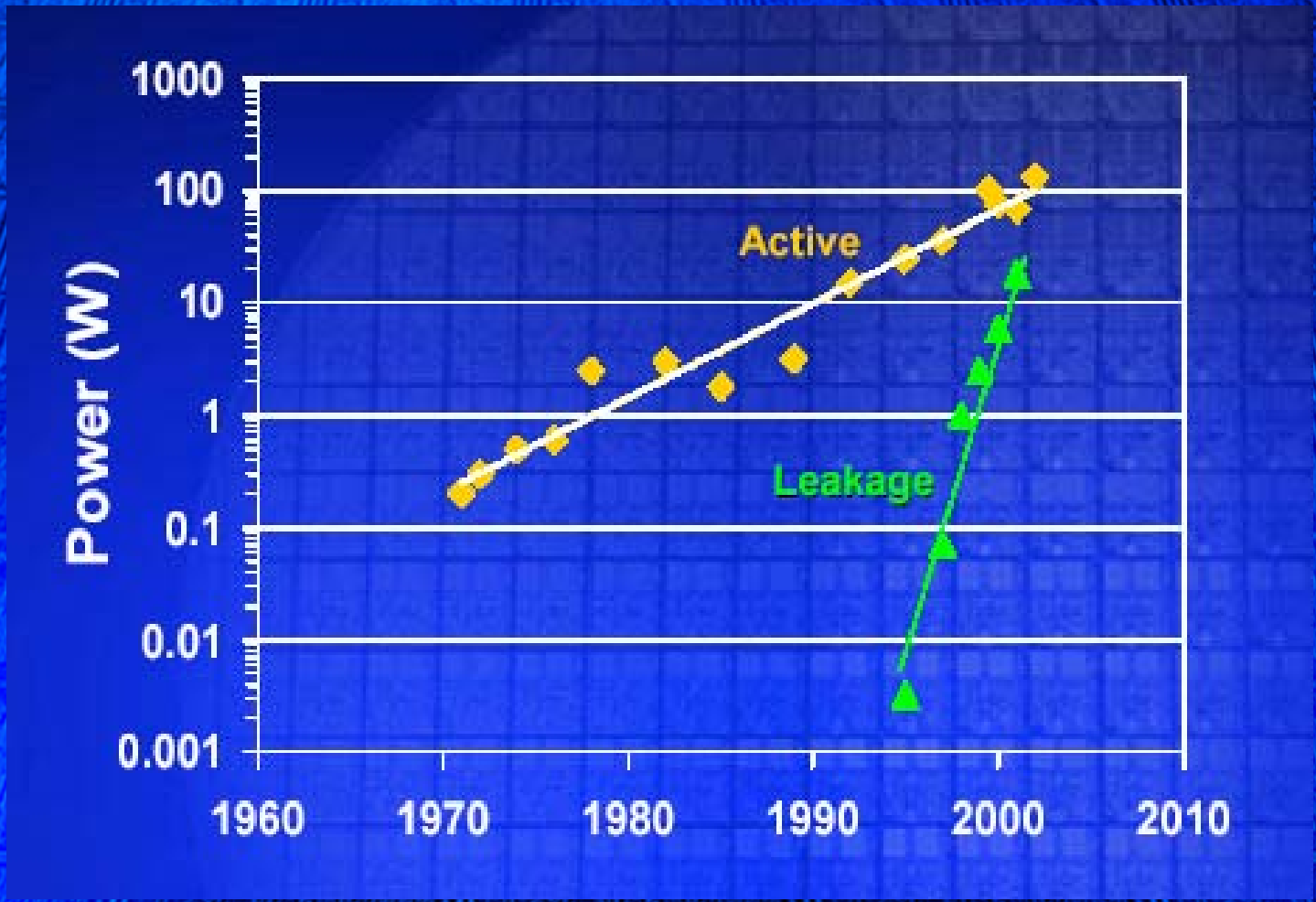
intel

Situational Analysis

- With Each Process Generation transistor density doubles
  - Frequency has increased by ~1.5X; ~1.3x in future
  - Vcc has scaled by about ~0.8x; ~0.9x in future
  - Capacitance has scaled by 0.7x
  - Total power may not scale down due to increased leakage
- Instruction Level Parallelism harder to find
- Increasing single-stream performance often requires non-linear increase in design complexity
- Many server applications are inherently parallel
- Parallelism exists in multimedia applications
- Multi-tasking usage models becoming popular

intel

Processor Power



intel

Design Complexity and Productivity factors

- Huge transistor budgets stress ability to design and verify complex chips
- Multi-core fits well with increasing transistor budgets
- Multi-core design addresses density/designer gap

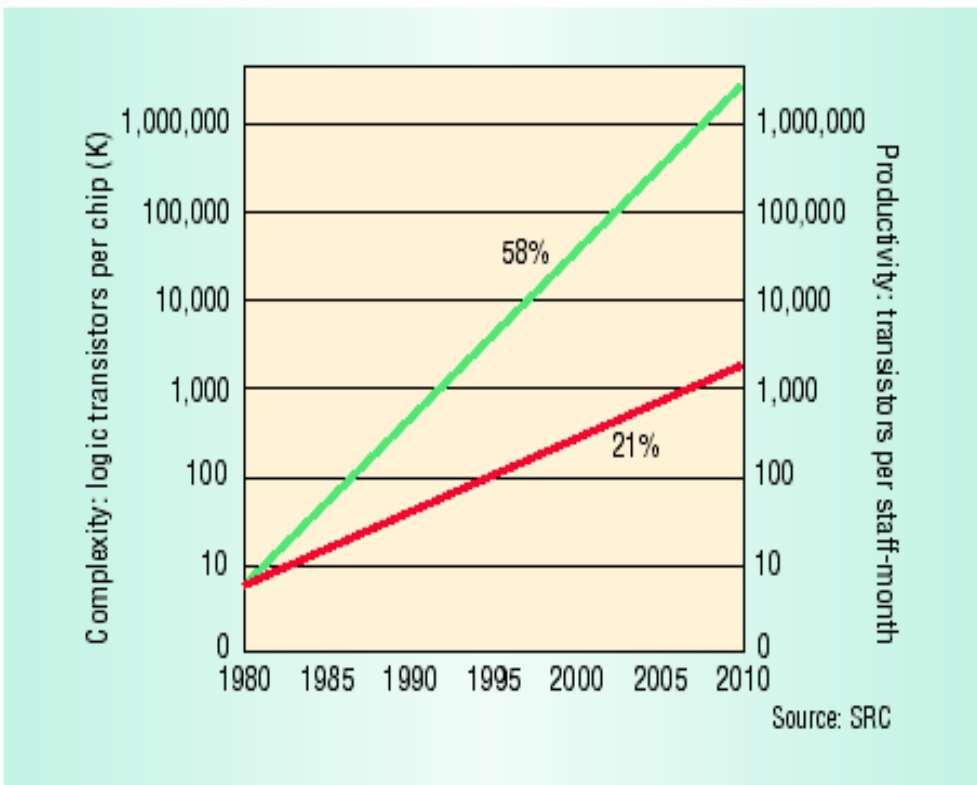


Figure 2. Design complexity and designer productivity. Since 1980, the design gap between growth in chip complexity and productivity growth in logic design tools has widened each year.

intel

Agenda

- Semiconductor Technology Evolution
- Design Challenges
- Why Multi-Core Processor Chips?
- Power/Performance Trade-Offs
- CMP Directions
- Beyond CMP
- Summary

©2005, Intel Corporation
Intel, the Intel logo, Pentium, Itanium and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries
*Other names and brands may be claimed as the property of others

Iron Law of Performance

- Execution Time is the product of
  - Path Length
  - Cycles Per Instruction (CPI)
  - Cycle Time
- CPI is the sum of
  - infinite-cache core cpi
  - miss rate * effective memory latency
- Bad (good) news is that performance does not scale up (down) linearly with frequency

intel

The Magic of Voltage Scaling

- Power = Capacitance * Voltage^2 * Frequency
- Frequency α Voltage in region of interest
- Power increases as the cube of Frequency
- Good news is that voltage scaling works
- 10% reduction in voltage yields
  - 10% reduction in frequency
  - 30% reduction in power
  - less than 10% reduction in performance

intel

Simple Dual Core Example

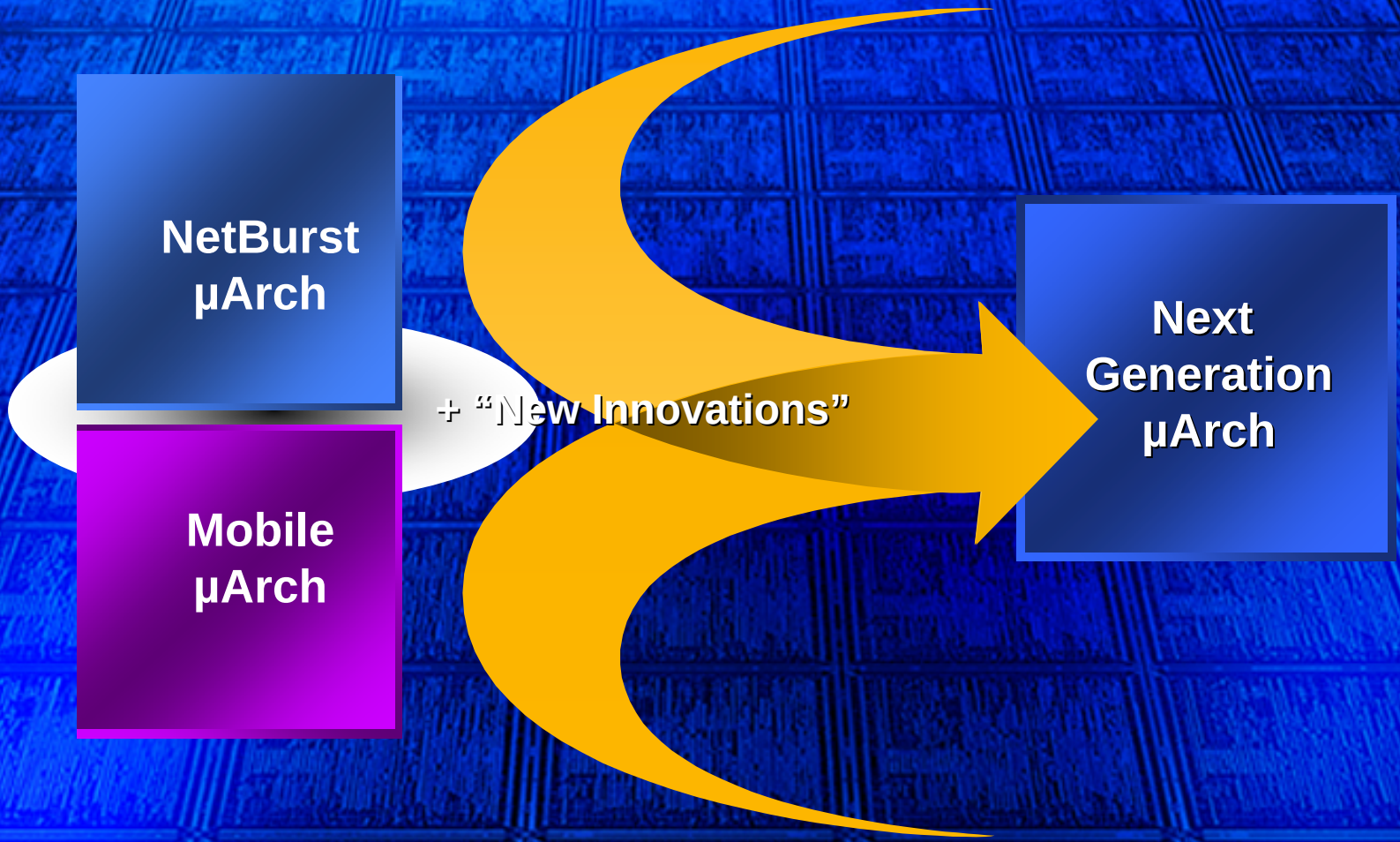
- Assume Single Core processor at 100W
  - 80W for core, 20W for cache and I/O
  - 50% die are is core
- Dual core within same power envelop
  - 20W for I/O and cache
  - 40W per core
  - Die size increases by 50%
  - Reduce voltage by 21% to reduce core power to 40W
  - Frequency reduces by ~20%
  - Single thread perf reduces by ~15%
  - Throughput increases by 70-80%

intel

Possible Improvements

- Develop new power efficient core
  - E.g. extensive clock gating
  - Big power savings with little or no performance loss
- Design a smaller core with lower performance
  - Area and power savings much greater than performance loss
  - Use larger number of cores
- Adjust frequency and power of each core with load factor
  - Inactive cores can be put in sleep mode
  - Maintain overall die power constant

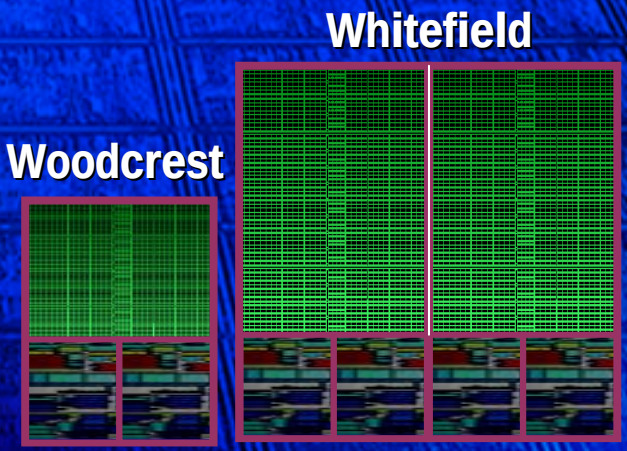
Intel's Next Generation Micro-Architecture



intel

Server Products

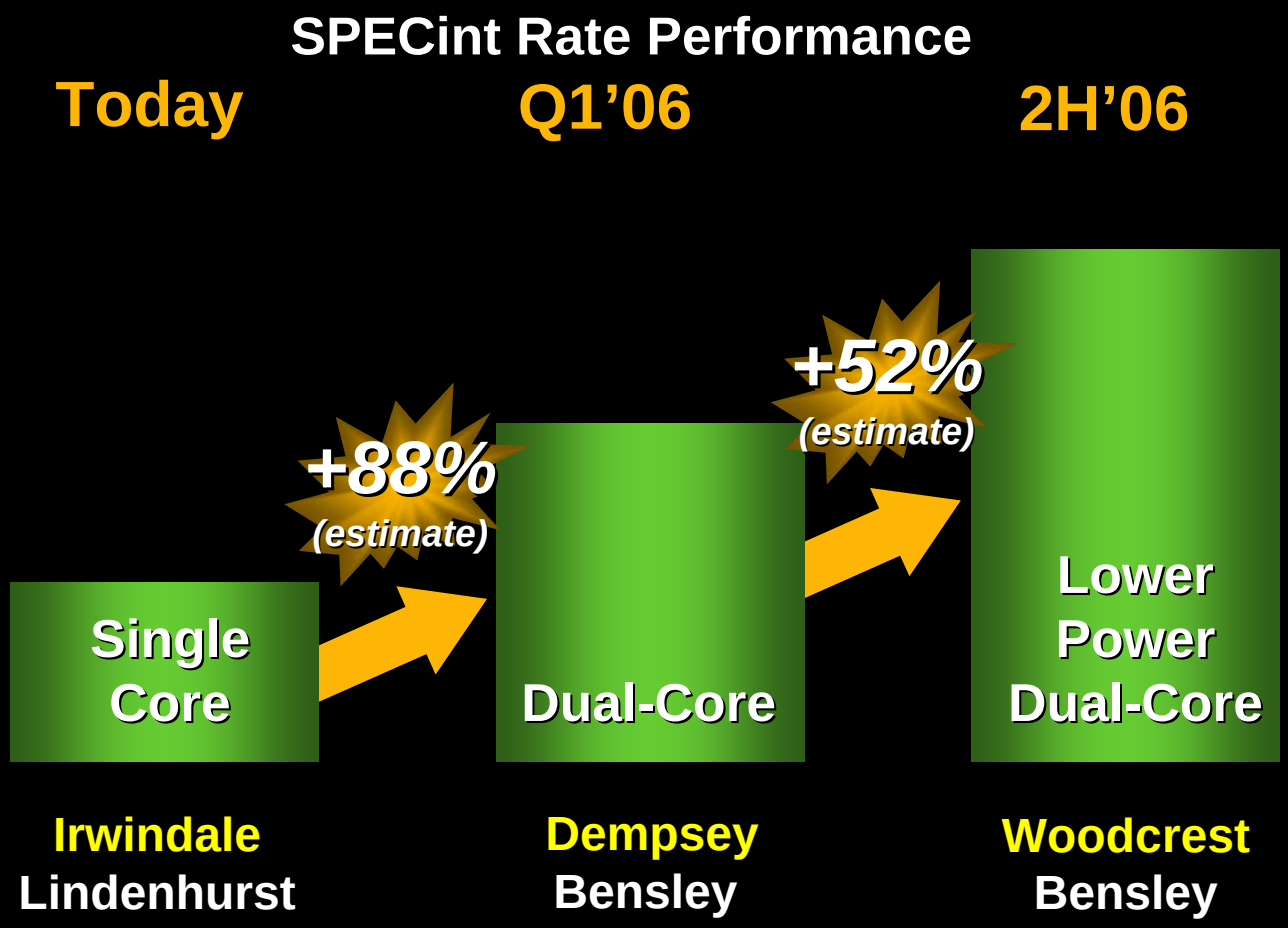
- 2 to 4 highly efficient cores
- Large/shared and scalable L2 cache
- 40% reduction in TDP power
- Server *Ts



- Greater compute density
- Lower energy consumption
- Greater application responsiveness
- Improved virtualization, manageability, and security

Enabling

Advancing Performance



Source: Intel Corporation. Estimated SPECint_rate2000 performance. Projections and technical specifications are based on internal analysis and subject to change

All dates and products specified are for planning purposes only and are subject to change.

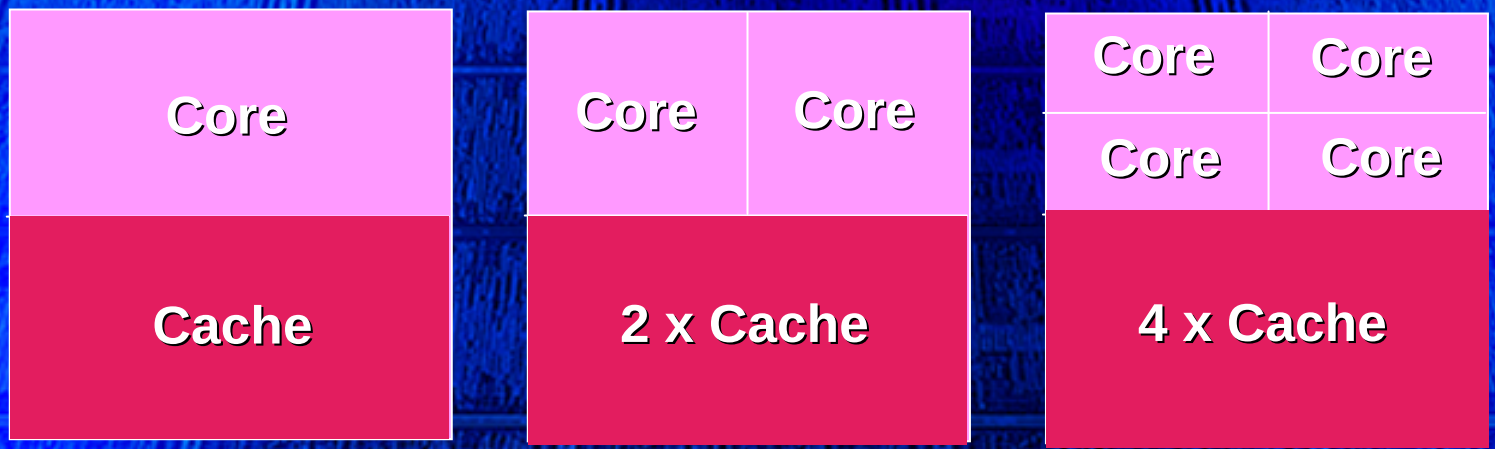
Agenda

- Semiconductor Technology Evolution
- Design Challenges
- Why Multi-Core Processor Chips?
- Power/Performance Trade-Offs
- CMP Directions
- Beyond CMP
- Summary

©2005, Intel Corporation
Intel, the Intel logo, Pentium, Itanium and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries
*Other names and brands may be claimed as the property of others

Possible Evolution

- Transistor density doubles with each process generation
- New generation enables complex new core
- Possible alternative design point
  - Double the cache capacity in same area
  - Double the number of processor cores
  - Frequency improves with process technology



Multi-Core Everywhere

	2005	2006	2007
SERVER	Shipping	>85%	~100%
DESKTOP PERFORMANCE	Shipping	>70%	>90%
MOBILE PERFORMANCE	Shipping	>70%	>90%

intel

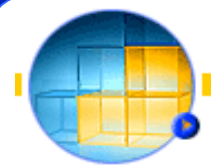
CMP Challenges

- How much Thread Level Parallelism is there in most workloads?
- Ability to generate code with lots of threads & performance scaling
- Thread synchronization
- Operating systems for parallel machines
- Single thread performance tradeoff
- Power limitations
- On-chip interconnect/cache infrastructure
- Memory and I/O bandwidth required

Intel's Software Tools and Support

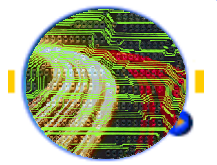
Thread Checker
Thread Profiler

Solutions, Blueprints,
Sizing/Scaling Guides



Math Kernel Libraries
Performance Primitives

Driver Optimization
Labs



Compilers

Solution Services
Developer Services

VTune™ Analyzers

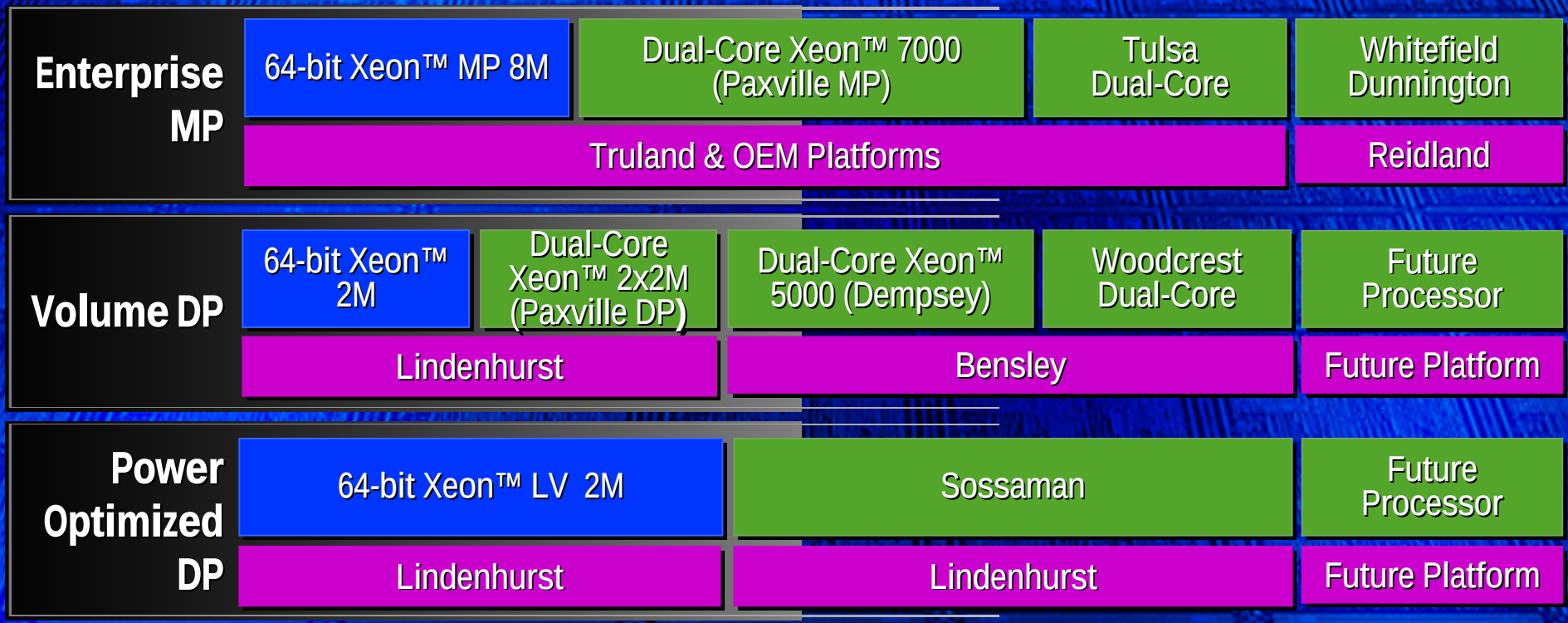
Software College
Early Access Programs

How Many Cores?

- Where does the doubling stop?
  - Driven by software issues
- Today Microsoft Windows supports only 64 threads!
- How many applications scale to 64 threads?
- How well does performance scale with thread count?

intel

Xeon Server & Workstation Roadmap



Agenda

- Semiconductor Technology Evolution
- Design Challenges
- Why Multi-Core Processor Chips?
- Power/Performance Trade-Offs
- CMP Directions
- Beyond CMP
- Summary

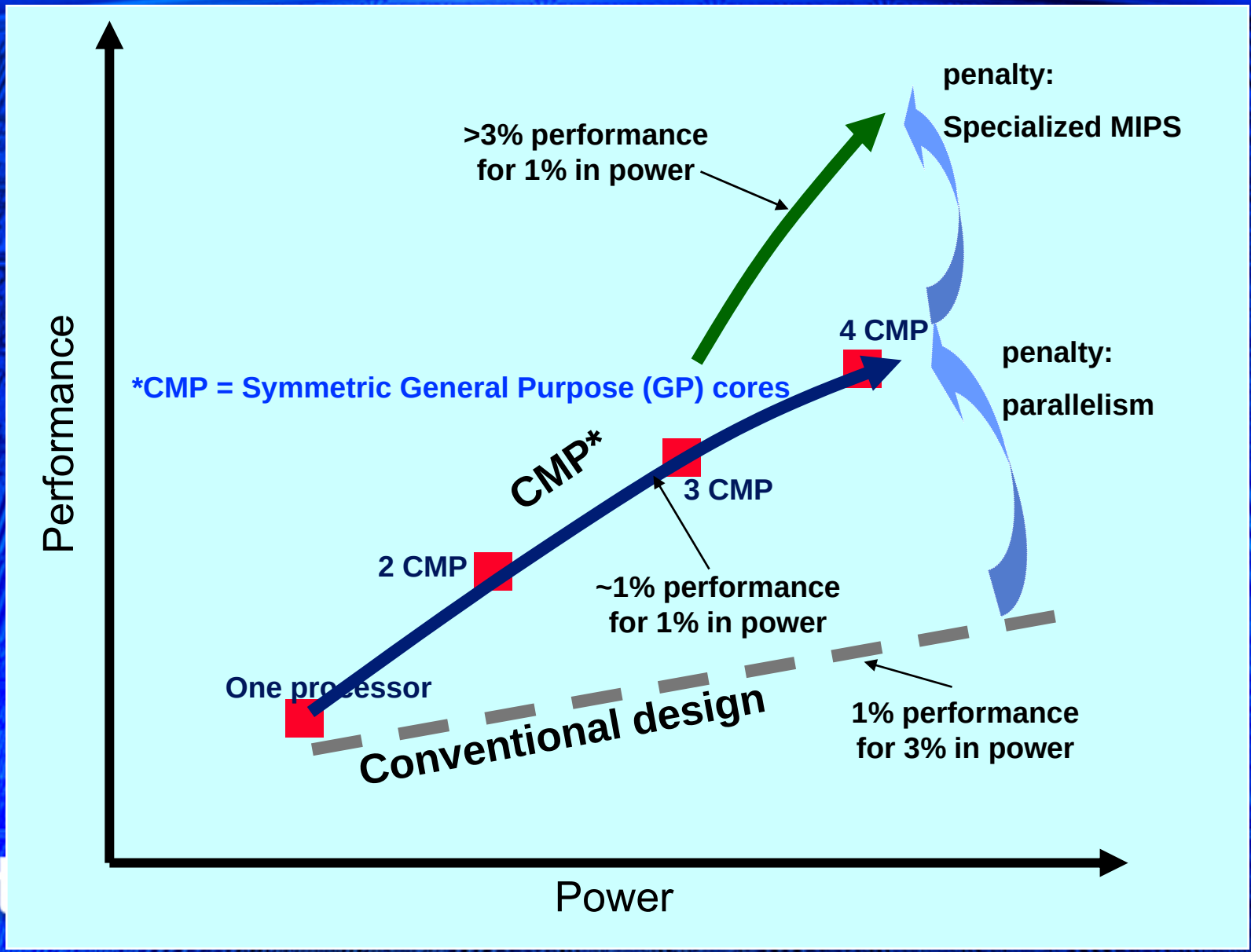
©2005, Intel Corporation
Intel, the Intel logo, Pentium, Itanium and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries
*Other names and brands may be claimed as the property of others

Looking Beyond CMP

- How far do we push the number of general purpose cores?
- Is there are role for application specific engines?
- Programming model for heterogeneous cores

intel

Improving Power Efficiency



Application Specific Engines

- Can achieve better power efficiency than general purpose cores
- Simpler design due to targeted application and lack of support for full operating system
- Challenge
  - Needs to support high volume application
  - Reconfigurable?
- Graphics and Multimedia engines are good candidates

Agenda

- Semiconductor Technology Evolution
- Design Challenges
- Why Multi-Core Processor Chips?
- Power/Performance Trade-Offs
- CMP Directions
- Beyond CMP
- Summary

©2005, Intel Corporation
Intel, the Intel logo, Pentium, Itanium and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries
*Other names and brands may be claimed as the property of others

Summary

- One billion transistors in 2005
- Chip Level Multiprocessing and large caches can exploit Moore's Law
- Amount of parallelism in future microprocessor systems will increase
- Heterogeneous cores may emerge eventually
- Need applications and tools that can exploit parallelism
- Design challenges and software issues remain

intel

Collaborate, Innovate, Lead!

Closing Thought

"Don't be encumbered by past history, go off and do something wonderful."

- Robert Noyce
Intel Co-founder